

# G4LDB: a database for discovering and studying G-quadruplex ligands

**Qian Li, Jun-Feng Xiang, Qian-Fan Yang, Hong-Xia Sun, Ai-Jiao Guan and Ya-Lin Tang\***

Beijing National Laboratory for Molecular Sciences (BNLMS), Center for Molecular Sciences, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, PR China

Received July 18, 2012; Revised October 11, 2012; Accepted October 21, 2012

## ABSTRACT

The G-quadruplex ligands database (G4LDB, <http://www.g4ldb.org>) provides a unique collection of reported G-quadruplex ligands to streamline ligand/drug discovery targeting G-quadruplexes. G-quadruplexes are guanine-rich nucleic acid sequences in human telomeres and gene promoter regions. There is a growing recognition for their profound roles in a wide spectrum of diseases, such as cancer, diabetes and cardiovascular disease. Ligands that affect the structure and activity of G-quadruplexes can shed light on the search for G-quadruplex-targeting drugs. Therefore, we built the G4LDB to (i) compile a data set covering various physical properties and 3D structure of G-quadruplex ligands; (ii) provide Web-based tools for G-quadruplex ligand design; and (iii) to facilitate the discovery of novel therapeutic and diagnostic agents targeting G-quadruplexes. G4LDB currently contains >800 G-quadruplex ligands with ~4000 activity records, which, to our knowledge, is the most extensive collection of its kind. It offers a user friendly interface that can meet a variety of data inquiries from researchers. For example, ligands can be searched for by name, molecular properties, structures, ligand activities and so on. Building on the reported data, the database also provides an online ligand design module that can predict ligand binding affinity in real time.

## INTRODUCTION

Tandem guanine-rich nucleic acid sequences, which are prevalent in human telomeres (1) and gene promoter regions (2,3), can form four stranded structures, called

G-quadruplexes (4). The stability of G-quadruplex structures potentially plays an important role in telomere maintenance (5,6) and in the regulation of gene expression (7,8). As such, G-quadruplexes can have profound implications in human diseases, such as cancer (9,10), diabetes (11,12) and cardiovascular disease (13). There is a growing interest in designing small molecule therapeutics that target G-quadruplexes (10,14). Ligands that selectively interact with promoter and telomere G-quadruplexes have the potential to regulate gene expression and the cell cycle and may, therefore, act as novel therapeutic and diagnostic agents (15,16). Many of the reported G-quadruplex ligands evaluated in biological assays have shown *in vitro* activities, including telomerase inhibition (17), oncogene downregulation (18) and suppression of cancer cell proliferation (19). Through xenograft models, a series of G-quadruplex stabilizers, such as BRACO-19 (20), TMPyP4 (21,22), RHPs4 (23,24) and telomestatin (25), showed *in vivo* anti-cancer activities associated with telomeres or *c-MYC*. Of particular note is quarfloxin (CX-3543), a G-quadruplex binder, which has shown great therapeutic potential and is currently undergoing phase II clinical trials for the treatment of low or intermediate grade neuroendocrine cancer (26).

In the past two decades, hundreds of small molecules with diverse chemical structures and physicochemical properties have been prepared and examined for their abilities to interact with G-quadruplexes (27,28). Concurrent to the search for G-quadruplex ligands is the development of our understanding of the essential moieties and binding modes involved in ligand/G-quadruplex interactions. For instance, early studies of G-quadruplex ligands mainly focused on compounds containing extended planar aromatic moieties, which form  $\pi$ - $\pi$  interactions with the terminal G-tetrad (29). Later studies of distamycin A (30) and an uncharged distamycin A analogue (31) explored a novel binding mode where two dimeric ligands bind to two opposite grooves of the

\*To whom correspondence should be addressed. Tel: +86 1 08 26 17 304; Fax: +86 1 06 25 22 090; Email: tangyl@iccas.ac.cn

© The Author(s) 2012. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial reuse, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

G-quadruplex. Furthermore, the discovery of novel G-quadruplex stabilizers, funtumine guanylhydrazone (32) and peimine (33), showed that non-planar molecules without aromatic moieties can also interact with G-quadruplexes through the groove binding mode. Thus, our understanding of ligand/G-quadruplex interactions continues to evolve. Meanwhile, researchers are actively exploring the vast chemical repertoire in search of chemical structures/motifs that have ultrahigh affinity towards G-quadruplexes. As systematic analysis of various G-quadruplex ligands may significantly facilitate understanding of the ligand/G-quadruplex interaction and promote discovery of novel G-quadruplex ligands, the demand for a comprehensive database of G-quadruplex ligands with cheminformatic tools is growing. The G4LDB, with systematically collected information of reported G-quadruplex ligands, provides a useful resource for the study of G-quadruplex ligands.

Computer aided drug design is a helpful tool in novel drug discovery and has been used to screen and design G-quadruplex ligands (19,33,34). Among molecular computational methods, molecular docking is one of the most important techniques, and it has been used to predict the interaction of small ligands and biological macromolecules (such as proteins and nucleic acids). As more and more ligand/G-quadruplex structures have been determined (35), a series of novel G-quadruplex ligands have been discovered using molecular docking (36–40). Aided by molecular docking techniques, costs associated with compound synthesis and/or activity evaluation have been significantly reduced. Despite these successful cases of G-quadruplex ligand discovery by molecular docking, wide and general use of this method remains limited because of the complicated docking assays. Performing molecular docking studies requires a strong background in molecular modelling, and most researchers, beyond the molecular modelling experts, often avoid using this technique to study G-quadruplexes. Easy-to-use docking tools that can extend the use of molecular docking to general G-quadruplex ligand researchers are still needed. With the online docking module, the G4LDB offers a convenient tool for ligand/G-quadruplex docking studies.

During the past decade, hundreds of G-quadruplex ligands have been discovered. Many of them show *in vitro* and *in vivo* activities, and some have entered clinical trials. Studies on G-quadruplex ligands are now among those at the forefront of drug discovery, and a database that includes comprehensive information of G-quadruplex ligands will greatly benefit such studies. Until now, there have been several G-quadruplex motif or sequence databases, the GRSDB (41), Greglist (42) and QuadBase (43). However, these databases focus on the motifs and sequences of G-quadruplexes rather than on G-quadruplex ligands. All of this motivated us to develop G4LDB, the first database to consist systematic information reported for G-quadruplex ligands and, moreover, to provide an online docking module to improve the efficiency of G-quadruplex ligand research and drug discovery.

## OVERVIEW OF THE DATABASE

The G4LDB (<http://www.g4ldb.org>) currently consists of two major components, the ligand database and the online prediction module. The ligand database currently contains >800 reported G-quadruplex ligands that are associated with 4000 related activity study records. This facilitates cheminformatic studies and systemic analysis of G-quadruplex ligands, and the search for novel binders. Each G-quadruplex ligand entry includes the molecular name, molecular weight, the molecular identifier (SMILES string), the number of H-bond donors and receptors, rotatable bonds, predicted *n*-octanol–water partition coefficient (logP) and G-quadruplex related activity reports (Figure 1D), such as the enhanced circular dichroism melting temperature, the inhibition of telomerase and the binding constant. The online docking module provides a Web-based interface for users to predict binding mode and affinities of newly designed compounds. There are 28 G-quadruplex targets, derived from all the published ligand/G-quadruplex complex structures in the RCSB Protein Data Bank (44), available for online docking through the web interface.

The G4LDB provides a user-friendly web interface with useful functions for retrieving and browsing the G-quadruplex ligand data set. The ‘Browse’ page (Figure 1E) allows the user to view the tabulated content of the database. The ‘Search’ page offers users three options for data querying. By using the ‘Quick Search’ function (Figure 1A), users can search ligands by the field, such as ligand name, target sequence, activity type, author name of ligand source literature and so on. The ‘Structure Search’ function (Figure 1B) implements a search of G-quadruplex ligands according to molecule or fragment structures. There are three options for structure-based searches, ‘Exact Match’, ‘Substructure Match’ and ‘Similarity Match’. The ‘Advanced Search’ function (Figure 1C) provides a series of data fields (molecular properties and activities) and can perform integrated information searches.

Another important part of G4LDB is the online prediction module, with which users can predict binding strength of their newly designed ligands. This module incorporated 28 docking models (Figure 2A) of G-quadruplexes, covering nearly all known binding modes reported in the RCSB protein data bank (such as the end stacking mode, intercalate mode, groove binding mode and loop binding mode). After designing the ligand and submitting the docking job (Figure 2C), system preparation related to molecular docking, binding evaluation and results extraction will be automatically carried out on the G4LDB server, and the results are sent to the user. A prediction of complex structure and binding affinity become available through a unique URL once the docking simulation finishes. The predicted binding modes and molecular interactions are rendered directly by the Jmol applet 3D visualization tools embedded in the web page (Figure 2D). The evaluation results, including the predicted binding affinities (in a text file) and ligand/G-quadruplex complex structure (in a Protein Data Bank file), are also downloadable from this page.

**A Quick search**

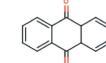
You can search by ligand ID (e.g. G4L0021), ligand name(e.g. TMPyP4), ligand activity or binding properties (e.g. Cytotoxicity), author name of ligand related literatures (e.g. Li)

**C Advanced Search**

Ligand Properties			
Name:	<input type="text"/>	Hydrogen Bond Acceptor:	Between <input type="text"/> and <input type="text"/>
Molecular Weight:	Between <input type="text"/> and <input type="text"/>	Hydrogen Bond Donor:	Between <input type="text"/> and <input type="text"/>
logP:	Between <input type="text"/> and <input type="text"/>	Rotatable Bonds:	Between <input type="text"/> and <input type="text"/>
Molecular Solubility:	Between <input type="text"/> and <input type="text"/>	Metal Ions:	Between <input type="text"/> and <input type="text"/>
Ligand Bibliographic Properties			
Author name:	<input type="text"/>		
Ligand Activities			
Activity Type:	Please select a activity type		Between <input type="text"/> and <input type="text"/>
<input type="button" value="Search and show ligands list"/>		<input type="button" value="Search and show activity list"/>	

**B Search by structure**

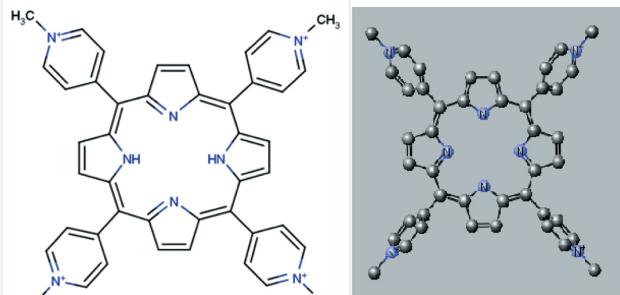
Draw a structure or open a file:



Smiles: O=C1C2=C=CC(=O)C2=C=CC1

Search Option:  Similarity Match  Similarity > 0.6  0-1  
 Exact Match  Substructure Match  Search

**D Ligand: TMPyP4 (tetra(Nmethyl- 4-pyridyl)porphine)**

  
[Download the molecule](#)

AlogP	9.656
AlogP98	7.045
Formal Charge	4
Formula	C <sub>45</sub> H <sub>44</sub> N <sub>4</sub>
Weight	678.826
Molecular Solubility	678.826
HBA Count	2
HBD Count	2
Rotatable Bonds	0
Metal Atoms	0
Smiles	<chem>C[n+]1ccccc1C2=Cc3c(cnc3C4=CC=CC=C4)C2Cc5ccccc5</chem>
Activity 1	
Ligand ID	G4L0073
Ligand Name	TMPyP4 (tetra(Nmethyl- 4-pyridyl)porphine)
Target ID	G4N0016
Target Sequence	TTTTGGGG
Activity 54	
Ligand ID	G4L0073
Ligand Name	TMPyP4 (tetra(Nmethyl- 4-pyridyl)porphine)
Target ID	G4N0016
Target Sequence	TTTTGGGG
Inter/Intra	inter
Activity	ITC-
Value Type	delta H
Value1 / Unit	-25.3 / kcal/mol(-1)
Value2 / Unit	2 / molar ratio
Reference	<a href="#">JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, 123(37), 5902-5913</a>

**E Browse G-quadruplex Ligands**

ID	Name	Structure	AlogP <sub>98</sub>	Formal Charge	Formula	M.W.	HBA Count	HBD Count	Rotatable Bonds	Metal Atoms
G4L0021	2,6-dimethoxyanthraquinone derivative compound 1		-0.339	2	C <sub>20</sub> H <sub>14</sub> O <sub>4</sub>	306.1752	4	4	10	0
G4L0022	Piperazine bis(1-(4-pyridinio)ethyl)-3,4,9,10-perylenetetracarboxylic dianide		1.905	2	C <sub>42</sub> H <sub>44</sub> N <sub>4</sub> O <sub>4</sub>	614.733	4	0	6	0
G4L0023	1,4-disubstituted amidonaphthalene-9,10-diones 9a		3.392	0	C <sub>20</sub> H <sub>14</sub> O <sub>4</sub>	516.631	4	2	8	0

Ligand ID	G4L0073	Methods	
Ligand Name	TMPyP4 (tetra(Nmethyl- 4-pyridyl)porphine)	A solution of DNA ([DNA] ) 0.25- 0.40 mM in bases) and H2TMPyP ([H2TMPyP] ) 3.0-5.0 mM was titrated into a solution of H2TMPyP at the same porphyrin concentration (i.e., at 3.0-5.0 mM) and the spectrum recorded from 400 to 500 nm after incubation for 5 min.	
Target ID	G4N0016		
Target Sequence	TTTTGGGG		
Inter/Intra	inter	Comments	
Activity	VIS-		
Value Type	delta lambda		
Value1 / Unit	10 / nm		
Value2 / Unit	0 /		
Reference	<a href="#">BIOCHEMISTRY, 37(9), 2709-2714</a>		

**A**

Step 1: Choose A Target Model

ID	Structure	PDB ID	PMID	G4 Form	Inter Intra	Active Site	Comment	Choose
1		1L1H	12547195	[d(GGGGTTTGGG)] <sub>2</sub>	inter		Activity Site 1	<a href="#">Choose</a>
2		1OKK	12670225	[d(TGGGTG)] <sub>4</sub>	inter		Activity Site 1	<a href="#">Choose</a>
3a		1N7M	14596797	[d(TTAGGGT)] <sub>4</sub>	inter		Activity Site 1	<a href="#">Choose</a>

< First | 1 | 2 | Next > | Last > | Go! | Total 2 pages)

**B**

PMID	12547195
Author	Heider SM, Parkington GH, Neidle S.
DOI	10.1093/nar/gss022-2836(02)01354-2
Source	J Mol Biol. 2003 Feb 7;329(1):117-25.
Sequence	[d(GGGGTTTGGG)] <sub>2</sub>
G4 Form	[d(GGGGTTTGGG)] <sub>2</sub>
Inter Intra	inter
Site Center	17.655 18.144 7.789
Site Dimension	15 15 15
Description	Crystal Structure of the Quadruplex-DNA-Drug Complex

[Compute binding free energy](#)

**C**

Step 2: Design an ligand

Draw a structure or open a file:

File Edit View Insert Atom Bond Structure Calculations Tools Help

87%

Smiles (required): CCCN(C1=C2C=C3C=CC4=CC(NCCSC)=CC5=CC6=C(C3=C45)C2=C

Job Description (required): G4LDB Binding Job

Email (optional):

Calculate

**D**

BFE Docking Results

Job Description	G4LDB Binding Job
PUB ID	2J17
Ligand SMILES	<chem>CCCN(C1=C2C=C3C=CC4=CC(NCCSC)=CC5=CC6=C(C3=C45)C2=C</chem>
pKa	5.18
Download	<a href="#">Download the result</a>

Display Mode 1 (Cartoon Mode)  
Display Mode 2 (Stick Mode)  
Display Mode 3 (Surface Mode. May take 30 seconds or more to generate molecular surface)  
Drag left mouse button to rotate; Scroll middle mouse button to zoom in/out

Jmol

**Figure 2.** Web interfaces of the online prediction module. (A) Choose target G-quadruplex structure model for evaluating the ligand/G-quadruplex interaction. (B) Detailed information of target G-quadruplex. (C) Design ligand for prediction. (D) The results page.

## MATERIALS AND METHODS

### Data acquisition

#### Structures of ligands

Since the first report of G-quadruplex ligands in the late 1980s, >2000 G-quadruplex reports (between 1988 and 2011) have been published. The ligands reported to interact with G-quadruplexes are included in

the database; their 2D chemical structures were constructed by Marvin Sketch (ChemAxon) directly from descriptions in the literature. The canonical SMILES strings of the ligands were generated by OpenBabel 2.3.0 (45). Other molecular descriptors, such as molecular weight, formal charge, AlogP (Ghose–Crippen–Viswanadhan octanol–water partition coefficient), molecular solubility, number of H-bond acceptors,

number of H-bond donors, number of metal atoms and number of rotatable bonds of the G-quadruplex ligands, were all computed using Discovery Studio 3.1 (Accelrys, San Diego, CA, USA).

#### **Data of activities**

The activity data of G-quadruplex ligands were collected mainly from the literature. Various indicators of activities have been recorded, such as inhibition of telomerase activity, inhibition of Taq polymerase activity, binding affinity of ligand/G-quadruplex. Besides quantitative activity indicators, brief descriptions of experimental methods by which the activity data were generated were also recorded to provide users with methodology information.

#### **Target information**

Structures and other information of target G-quadruplexes were mainly obtained from the RCSB Protein Data Bank. Currently, 28 ligand/G-quadruplex complex structures are included in the RCSB Protein Data Bank. All of these structures were processed and prepared to construct G-quadruplex target models.

#### **System data updates**

New data related to G-quadruplex ligands will be included in the database after twice yearly bibliographic reviews (updates every April and October, approximately).

#### **Structure searching tools for G-quadruplex ligands**

In the structure search function, molecular structures can be directly constructed by the graphical user interface of MarvinSketch 5.10.0 (ChemAxon) embedded in the structure searching web page. Based on OpenBabel 2.3.0, three modes of structure searching, the exact match, substructure match and similarity match, are provided by the searching function. In the exact match mode, canonical SMILES matching only returns the structure identical to the query molecule. In the substructure match mode, however, a molecular search based on fragment SMARTS matching will list all ligands containing the queried molecular fragment. The molecular similarity between the user-specified structure and ligands in the G4LDB is quantified by the Tanimoto similarity coefficient, based on the linear fingerprint of fragments indices. The Tanimoto similarity coefficient, ranging from 0 to 1, reflects the level of similarity between two molecules; 0 indicates minimal similarity, and 1 indicates the maximum. In the structure similarity searching mode, users can specify molecular structure and similarity threshold. The structure similarity search will automatically calculate and sort the Tanimoto similarity coefficient for each ligand in G4LDB and query structures and will only return similarity coefficient values higher than the threshold specified by the user.

#### **Molecular visualization**

G4LDB can display a series of molecular visualization effects; the 2D structure for each ligand is displayed as an image generated by MarvinSketch (ChemAxon);

3D structures of ligands are represented using the MarvinView (ChemAxon) applet embedded in these web pages. Online computational results of predicted molecular interactions between a small molecule and a G-quadruplex can be displayed in various styles, for example, cartoon models, all atom ball-and-stick models and molecular surface models, using the Jmol applet.

#### **Online docking**

To prepare docking models of G-quadruplexes, all 3D structures of ligand/G-quadruplex complexes recorded in the RCSB Protein Data Bank (from January 1994 to June 2012) were retrieved. For each retrieved ligand/G-quadruplex complex, a grid box ( $15\text{ \AA} \times 15\text{ \AA} \times 15\text{ \AA}$ ), centred at the geometrical centre of the ligand and containing grid points with  $0.375\text{ \AA}$  spacing, was prepared (Supplementary Table S1). The ligand was then removed from the structure to leave an empty binding pocket, defined as the grid box. Detailed information on the definition of the active site for each model is shown in the Supplementary Data.

The G4LDB online docking proceeds in two steps. The first step is small ligand preparation. The 2D chemical structure constructed in the ligand design page by the user is converted to a 3D structure. The 3D ligand structure then undergoes geometry optimization to produce the initial ligand structure. The 2D to 3D transformation, geometry optimization and molecular format transformation are performed by OpenBabel 2.3.0 and AutoDockTools-1.5.4 (46) at the server end.

The second step is molecular docking. Grid maps of the pre-defined active site for each atom type present in the ligand are calculated by AutoGrid (47) before docking a ligand to the target G-quadruplex. After assigning Gasteiger partial charges to the atoms in the ligand, the Lamarckian genetic algorithm based molecular docking and calculation of ligand/G-quadruplex dissociation constant are carried out by AutoDock (47). The docking parameters are set as follows: the ligand translation step is set to  $2.0\text{ \AA}$ , the ligand quaternion and torsion step are both set to  $50^\circ$ , the maximum number of energy evaluations is set to  $1.0 \times 10^6$ , the maximum number of genetic algorithm operations is set to  $2.7 \times 10^4$ , the number of individuals in the population is set to 150, the rate of mutation and crossover are set to 0.02 and 0.8, respectively. Other parameters are all set as default. When searching the conformational and orientational spaces of the ligand with rotatable bonds having full flexibility, the structure of the G-quadruplex is kept rigid. For each docking evaluation, 20 independent runs are performed to evaluate different ligand poses, and only the most favourable pose is dumped to the results file.

#### **Database server implementation**

The G4LDB is installed on CentOS 5.3 server workstations. An Apache 2.2.3 server is used as the web server platform, and the website was built with PHP 5.2. The multiple types of data are organized, managed and stored using the MySQL 5.1 relational database management system. To construct an easily accessible searching

and retrieval system, the graphic chemical editor MarvinSketch (ChemAxon) was used to build interactive web interfaces of the G4LDB. The substructure matching and molecular similarity prediction are accomplished by OpenBabel 2.3.0. Jmol applets are embedded in the interface to render 3D structures of small ligands and of predicted complex structures. The G4LDB site is best viewed by Firefox, Safari, Google Chrome and Opera in which JavaScript (version >1.7) and Java (version >6) are enabled. The platforms that have been tested and that show stability with G4LDB are listed at <http://www.g4ldb.org/ci2/index.php/other/compatibility>.

## DISCUSSION

Integrated with a series of useful functions, the G4LDB provides a unique platform for the design and discovery of G-quadruplex targeting ligands/drugs. In the examples below we illustrate potential applications of the G4LDB.

### Browsing G-quadruplex ligands and checking information for a specified ligand

More than 800 G-quadruplex ligands are currently included in the G4LDB. By single clicking the ‘Browse → G-quadruplex ligands’ button, users will find tables listing the G-quadruplex ligands that compose the G4LDB (Figure 1E). Shown in the table is brief information for each ligand, whereas detailed information can be seen by simply clicking the ID link or the 2D structure of the ligand from the table list. General information along with different physicochemical properties and activity information are shown in the ‘ligand details’ page (Figure 1D). 2D and 3D structures of each ligand are also displayed on this page and can be downloaded in mol or mol2 formats. By reading the detailed information of the G-quadruplex ligand, users can understand the ligand more comprehensively. For example, by reading the ligand details page for TMPyP4, users can access not only the structure and physicochemical properties of the ligand but also >50 records of activities reported in the literature, for example, molecular binding affinity, *in vitro* inhibition of different cell lines and *in vivo* data on different animal models. From another point of view, this single web page represents the development status of TMPyP4, which provides an important reference for developing other G-quadruplex targeted drugs or probes.

### Designing small ligands and predicting molecular interactions between small molecules and G-quadruplexes

#### *Two-step job submission*

In the ‘Compute → Online Docking’ module of G4LDB, prediction of molecular interactions between a small molecule and G-quadruplexes can be achieved by two simple steps.

The first step is to select a target G-quadruplex model of interest. There are currently 28 G-quadruplex structures available as target models, all of which are listed in the ‘Step 1 Target Selection’ page (Figure 2A). These computational target models are derived from the 28 currently resolved ligand/G-quadruplex 3D structures. This page

displays brief information of each target, such as introduction of the structure, the sequence forming the structure, the chain type of the G-quadruplex, the Protein Data Bank ID and the reference link to the literature that reported the structure. Detailed information (Figure 2B) of the target can be retrieved by clicking the PDB\_ID or structure image of the target. Once the user identifies the target model of interest, target selection procedure can be accomplished by clicking the ‘Choose’ button. The computing system will automatically define the selected target G-quadruplex structure model and then take the user to the second step.

The second step is to design the small ligand. After choosing the target G-quadruplex model, the user is taken to the ligand designing page (Figure 2C), where a small ligand can be constructed in various ways. The simplest way is to build small ligands directly in the online molecular designing module (Marvin Sketch). In this way, users can design any small molecule. If small molecules have been designed previously by other software packages and saved as files, they can also be loaded by the designing module. By using the ‘File → Open...’ function in the menu, users can load external (local) ligand files in various formats, such as the Marvin document format (\*.mrw), the MDL mol format and sdf format (\*.mol, \*.sdf), the Symyx SKC format (\*.skc), the ChemDraw cdx format (\*.cdx), the Tripos mol2 format (\*.mol2), the Protein Data Bank format (\*.pdb) and the Gaussian output format (\*.gout). Once the small molecule is constructed or loaded, the SMILES string of the molecule appears in the box under the designing module. The user can then assign a short job name for the online prediction work and click the ‘calculate’ button. The online prediction work can be tracked by a dedicated URL provided on the confirmation page after submission. If the user’s email address has been submitted, the molecular prediction results and a link to the web page displaying the results are also sent to the user through an email when the prediction calculation is completed.

#### *Result retrieval*

After successful job submission, it usually takes 15–20 min for the G4LDB server to evaluate the ligand/G-quadruplex interactions. The results are available through an automatically generated URL on the confirmation page. The predicted binding modes and molecular interactions are rendered directly by the Jmol applet 3D visualization tools embedded in the web page (Figure 2D). Users can drag the left mouse key to rotate the 3D models of ligand/G-quadruplexes. Similarly, users can also drag the middle mouse key to zoom in and out of the models. Moreover, users can simply switch among different 3D display modes, such as the cartoon style, the stick mode and the molecular surface style, by clicking the option buttons in the result page. The evaluation results zipfile, consisting of the predicted ligand/G-quadruplex  $pK_i$  (a text file) and ligand/G-quadruplex complex structure (a pdb file), is also downloadable from this page.

Current version of the G4LDB user interface only supports single ligand submission for binding mode/affinity prediction. We plan to address this limitation by

introducing batch submission functions in future versions of the G4LDB. Also worthy of note is the inherent limitation of molecular docking and empirical force fields, the G4LDB online docking module has difficulty in handling organometallic compounds. Thus, we do not recommend the evaluation of binding affinity for metallo-ligands with the current version of the G4LDB docking module. We plan to address this issue in the near future by introducing computational tools with higher levels of theory into the prediction module.

### Filtering drug-like G-quadruplex ligands by their chemical descriptors

G4LDB provides a series of cheminformatic tools to facilitate the search of ligands with desired properties, such as 'drug-likeness'. Lipinski's Rule of Five (48) is a popular empirical measure of 'drug-likeness' for a chemical compound based on physicochemical properties. According to the Rule of Five, an active compound with good bioavailability should have a molecular weight <500, an octanol–water partition coefficient ( $\log P$ ) <5, hydrogen-bond donors <5 and hydrogen-bond acceptors <10. By filtering G-quadruplex ligands by the Rule of Five, G4LDB ligands that possess drug-like properties will be returned, and these ligands may be considered as drug candidates for oral delivery.

### Building quantitative structure-activity relationship models based on selected activity data

The quantitative structure-activity relationship (QSAR) approach provides a useful method to guide the design or optimization of ligands with respect to certain physicochemical or pharmacological properties. The G4LDB serves as an ideal source of both structural and activity data, which are the most essential factors for building predictive QSAR models.

Here, we take structural stabilization capabilities measured by fluorescence resonance energy transfer (FRET) as an example. First, users can access the advanced search page by selecting the 'Search → Advanced Search' option on the main menu. Then, by selecting 'FRET' from the drop-down list, specifying the value range (optional) and pressing the 'Search and show activity list' button, users can obtain a table that lists selected ligand structures and relative melting temperatures. In this way, users can use G-quadruplex ligand structures and quantitative activity values to perform QSAR studies.

### CONCLUSIONS

The G4LDB is a unique platform that contains a collection of >800 reported G-quadruplex ligands and an online ligand design module that can predict ligand binding affinity in real time. We built the G4LDB to (i) compile a comprehensive data set covering various physical properties of G-quadruplex ligands; (ii) provide Web-based tools for G-quadruplex ligand design; and (iii) facilitate the discovery of novel therapeutic and diagnostic agents targeting G-quadruplexes. We believe that the

development and operation of this database will significantly facilitate the discovery of novel drugs and diagnostics targeting G-quadruplexes.

To further improve the database, we plan to incorporate the following data sets and functionalities in future versions of the G4LDB. More organized data related to G-quadruplex ligands will be added to the G4LDB, such as spectroscopic data of ligands and detailed information of experimental methods. Emerging studies suggest that RNA G-quadruplexes exist in the 5'-UTRs of many genes of clinical interest, and because such structural elements can influence translation (49), the ligands of RNA G-quadruplexes will also be incorporated in the near future. To screen and design ligands of unsolved G-quadruplex structures, homology modelling methods may also be applied to extend the target G-quadruplex structure and, therefore, provide more screening models.

G-quadruplex ligand research is an emerging field in the discovery of novel drugs and diagnostics. We believe the G4LDB will provide researchers with a valuable resource for discovering and designing novel G-quadruplexes.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1.

### ACKNOWLEDGEMENTS

The authors thank the National Scientific Computing Grid (SCGrid) and the Supercomputer Center of the Chinese Academy of Sciences (SCCAS) for computational facilities.

### FUNDING

National Natural Science Foundation of China [91027033, 81072576, 21205121 and 31200576]; Key Program of the Chinese Academy of Sciences [KJCX2-EW-N06-01]. Funding for open access charge: National Natural Science Foundation of China [91027033].

*Conflict of interest statement.* None declared.

### REFERENCES

1. Sundquist,W.I. and Klug,A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
2. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
3. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
4. Frankkamenetski,M. (1992) The quadruplex and the vase. *Nature*, **356**, 105–105.
5. Sun,D., Thompson,B., Cathers,B.E., Salazar,M., Kerwin,S.M., Trent,J.O., Jenkins,T.C., Neidle,S. and Hurley,L.H. (1997) Inhibition of human telomerase by a G-quadruplex-interacting compound. *J. Med. Chem.*, **40**, 2113–2116.

6. Xu,Y. (2011) Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA. *Chem. Soc. Rev.*, **40**, 2719–2740.
7. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
8. Rodriguez,R., Miller,K.M., Forment,J.V., Bradshaw,C.R., Nikan,M., Britton,S., Oelschlaegel,T., Xhemalce,B., Balasubramanian,S. and Jackson,S.P. (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.*, **8**, 301–310.
9. De Cian,A., Lacroix,L., Douarre,C., Temime-Smaali,N., Trentesaux,C., Riou,J.F. and Mergny,J.L. (2008) Targeting telomeres and telomerase. *Biochimie*, **90**, 131–155.
10. Balasubramanian,S., Hurley,L.H. and Neidle,S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261–275.
11. Schonhoff,J.D., Bajracharya,R., Dhakal,S., Yu,Z.B., Mao,H.B. and Basu,S. (2009) Direct experimental evidence for quadruplex-quadruplex interaction within the human ILPR. *Nucleic Acids Res.*, **37**, 3310–3320.
12. Connor,A.C., Frederick,K.A., Morgan,E.J. and McGown,L.B. (2006) Insulin capture by an insulin-linked polymorphic region G-quadruplex DNA oligonucleotide. *J. Am. Chem. Soc.*, **128**, 4986–4991.
13. Zhou,W.H., Brand,N.J. and Ying,L.M. (2011) G-quadruplexes—novel mediators of gene function. *J. Cardiovasc. Transl. Res.*, **4**, 256–270.
14. Neidle,S. (2010) Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.*, **277**, 1118–1125.
15. Folini,M., Venturini,L., Cimino-Reale,G. and Zaffaroni,N. (2011) Telomeres as targets for anticancer therapies. *Expert Opin. Ther. Tar.*, **15**, 579–593.
16. Collie,G.W. and Parkinson,G.N. (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.*, **40**, 5867–5892.
17. Patel,D.J., Phan,A.T. and Kuryavyi,V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.*, **35**, 7429–7455.
18. Huppert,J.L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.*, **37**, 1375–1384.
19. Cuesta,J., Read,M.A. and Neidle,S. (2003) The design of G-quadruplex ligands as telomerase inhibitors. *Mini. Rev. Med. Chem.*, **3**, 11–21.
20. Burger,A.M., Dai,F., Schultes,C.M., Reszka,A.P., Moore,M.J., Double,J.A. and Neidle,S. (2005) The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res.*, **65**, 1489–1496.
21. Thakur,R.K., Kumar,P., Halder,K., Verma,A., Kar,A., Parent,J.L., Basundra,R., Kumar,A. and Chowdhury,S. (2009) Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclelease hypersensitive element induces c-MYC expression. *Nucleic Acids Res.*, **37**, 172–183.
22. Ofer,N., Weisman-Shomer,P., Shklover,J. and Fry,M. (2009) The quadruplex r(CGG)n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res.*, **37**, 2712–2722.
23. Cookson,J.C., Dai,F., Smith,V., Heald,R.A., Laughton,C.A., Stevens,M.F. and Burger,A.M. (2005) Pharmacodynamics of the G-quadruplex-stabilizing telomerase inhibitor 3,11-difluoro-6,8,13-trimethyl-8H-quin[4,3,2-k]acridinium methosulfate (RHPS4) *in vitro*: activity in human tumor cells correlates with telomere length and can be enhanced, or antagonized, with cytotoxic agents. *Mol. Pharmacol.*, **68**, 1551–1558.
24. Salvati,E., Leonetti,C., Rizzo,A., Scarsella,M., Mottolese,M., Galati,R., Sperduti,I., Stevens,M.F., D'Incàrci,M., Blasco,M. et al. (2007) Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J. Clin. Invest.*, **117**, 3236–3247.
25. Tauchi,T., Shin-ya,K., Sashida,G., Sumi,M., Okabe,S., Ohyashiki,J.H. and Ohyashiki,K. (2006) Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: *in vitro* and *in vivo* studies in acute leukemia. *Oncogene*, **25**, 5719–5725.
26. Drygin,D., Siddiqui-Jain,A., O'Brien,S., Schwaebe,M., Lin,A., Bliesath,J., Ho,C.B., Proffitt,C., Trent,K., Whitten,J.P. et al. (2009) Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res.*, **69**, 7653–7661.
27. Georgiades,S.N., Abd Karim,N.H., Suntharalingam,K. and Vilar,R. (2010) Interaction of metal complexes with G-quadruplex DNA. *Angew. Chem. Int. Ed. Engl.*, **49**, 4020–4034.
28. Monchaud,D. and Teulade-Fichou,M.P. (2008) A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.*, **6**, 627–636.
29. Read,M.A., Wood,A.A., Harrison,J.R., Gowan,S.M., Kelland,L.R., Dosanjh,H.S. and Neidle,S. (1999) Molecular modeling studies on G-quadruplex complexes of telomerase inhibitors: structure-activity relationships. *J. Med. Chem.*, **42**, 4538–4546.
30. Martino,L., Virno,A., Pagano,B., Virgilio,A., Di Micco,S., Galeone,A., Giancola,C., Bifulco,G., Mayol,L. and Randazzo,A. (2007) Structural and thermodynamic studies of the interaction of distamycin a with the parallel quadruplex structure [d(TGGGGT)](4). *J. Am. Chem. Soc.*, **129**, 16048–16056.
31. Cosconati,S., Marinelli,L., Trotta,R., Virno,A., De Tito,S., Romagnoli,R., Pagano,B., Limongelli,V., Giancola,C., Baraldi,P.G. et al. (2010) Structural and conformational requisites in DNA quadruplex groove binding: another piece to the puzzle. *J. Am. Chem. Soc.*, **132**, 6425–6433.
32. Brassart,B., Gomez,D., De Cian,A., Paterski,R., Montagnac,A., Qui,K.H., Temime-Smaali,N., Trentesaux,C., Mergny,J.L., Gueritte,F.O. et al. (2007) A new steroid derivative stabilizes g-quadruplexes and induces telomere uncapping in human tumor cells. *Mol. Pharmacol.*, **72**, 631–640.
33. Li,Q., Xiang,J.F., Li,X.D., Chen,L.R., Xu,X.J., Tang,Y.L., Zhou,Q.J., Li,L., Zhang,H., Sun,H.X. et al. (2009) Stabilizing parallel G-quadruplex DNA by a new class of ligands: two non-planar alkaloids through interaction in lateral grooves. *Biochimie*, **91**, 811–819.
34. Alcaro,S., Artese,A., Costa,G., Distinto,S., Ortuso,F. and Parrotta,L. (2011) Conformational studies and solvent-accessible surface area analysis of known selective DNA G-quadruplex binders. *Biochimie*, **93**, 1267–1274.
35. Neidle,S. and Parkinson,G.N. (2008) Quadruplex DNA crystal structures and drug design. *Biochimie*, **90**, 1184–1196.
36. Cosconati,S., Marinelli,L., Trotta,R., Virno,A., Mayol,L., Novellino,E., Olson,A.J. and Randazzo,A. (2009) Tandem application of virtual screening and NMR experiments in the discovery of brand new DNA quadruplex groove binders. *J. Am. Chem. Soc.*, **131**, 16336–16337.
37. Alcaro,S., Costa,G., Distinto,S., Moraca,F., Ortuso,F., Parrotta,L. and Artese,A. (2012) The polymorphisms of DNA G-quadruplex investigated by docking experiments with telomestatin enantiomers. *Curr. Pharm. Des.*, **18**, 1873–1879.
38. Haider,S. and Neidle,S. (2010) Molecular modeling and simulation of G-quadruplexes and quadruplex-ligand complexes. *Methods Mol. Biol.*, **608**, 17–37.
39. Redman,J.E., Granadino-Roldan,J.M., Schouten,J.A., Ladame,S., Reszka,A.P., Neidle,S. and Balasubramanian,S. (2009) Recognition and discrimination of DNA quadruplexes by acridine-peptide conjugates. *Org. Biomol. Chem.*, **7**, 76–84.
40. Kim,M.Y., Vankayalapati,H., Shin-Ya,K., Wierzba,K. and Hurley,L.H. (2002) Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex. *J. Am. Chem. Soc.*, **124**, 2098–2099.
41. Kikin,O., Zappala,Z., D'Antonio,L. and Bagga,P.S. (2008) GRSDB2 and GRS\_UTRdb: databases of quadruplex forming G-rich sequences in pre-mRNAs and mRNAs. *Nucleic Acids Res.*, **36**, D141–D148.
42. Zhang,R., Lin,Y. and Zhang,C.T. (2008) Greglist: a database listing potential G-quadruplex regulated genes. *Nucleic Acids Res.*, **36**, D372–D376.
43. Yadav,V.K., Abraham,J.K., Mani,P., Kulshrestha,R. and Chowdhury,S. (2008) QuadBase: genome-wide database of G4 DNA—occurrence and conservation in human, chimpanzee,

- mouse and rat promoters and 146 microbes. *Nucleic Acids Res.*, **36**, D381–D385.
44. Rose,P.W., Beran,B., Bi,C., Bluhm,W.F., Dimitropoulos,D., Goodsell,D.S., Prlic,A., Quesada,M., Quinn,G.B., Westbrook,J.D. *et al.* (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.*, **39**, D392–D401.
45. O'Boyle,N.M., Banck,M., James,C.A., Morley,C., Vandermeersch,T. and Hutchison,G.R. (2011) Open Babel: an open chemical toolbox. *J. Cheminform.*, **3**, 33.
46. Sanner,M.F. (1999) Python: a programming language for software integration and development. *J. Mol. Graph. Model.*, **17**, 57–61.
47. Morris,G.M., Huey,R., Lindstrom,W., Sanner,M.F., Belew,R.K., Goodsell,D.S. and Olson,A.J. (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.*, **30**, 2785–2791.
48. Lipinski,C.A., Lombardo,F., Dominy,B.W. and Feeney,P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, **46**, 3–26.
49. Bugaut,A. and Balasubramanian,S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40**, 4727–4741.